Historic, archived document

Do not assume content reflects current scientific knowledge, policies, or practices.

SHOUT
SHARE
IT

UNITED STATES DEPARTMENT OF AGRICULTURE

BULLETIN No. 504

OFFICE OF THE SECRETARY
Contribution from the Office of Farm Management
W. J. SPILLMAN, Chief

Washington, D. C. PROFESSIONAL PAPER May 23, 1917

THE THEORY OF CORRELATION AS APPLIED TO FARM-SURVEY DATA ON FATTENING BABY BEEF.

By H. R. TOLLEY, *Scientific Assistant.*

CONTENTS.

	Page.		Page.
Introduction	1	Interpretation of the coefficients	8
Theory of correlation	1	Summary	13
Computation of the coefficients	6		

INTRODUCTION.

This paper sets forth the results of an experiment in applying the theory of correlation, hitherto used chiefly in the analysis of biological, sociological, psychological, and meteorological statistics,[1] to the study of some of the data of the Office of Farm Management.

The material for the investigation was obtained from 67 records taken during the years 1914 and 1915 from farmers of the corn belt who were fattening baby beef for market.[2] The factors considered were: The profit or loss per head, the weight, value per hundredweight, value of feed consumed per head, cost at weaning time, and date of sale (see Table I). Coefficients of correlation were computed for every pair of these factors and used as a measure of the relationship existing between them.

THEORY OF CORRELATION.

The writer will not attempt a detailed explanation of the theory of correlation but will discuss briefly the meaning of coefficients of correlation and the method by which they are obtained.

[1] Yule, G. U.: "Introduction to the Theory of Statistics," 1912. Yule, G. U.: "On the Theory of Correlation," Jour. Roy. Stat. Soc., 1897, p. 812. Davenport, C. B.: "Statistical Methods, With Special Reference to Biological Variation," 1914. Hooker, R. H.: "The Correlation of the Weather and the Crops," Jour. Roy. Stat. Soc., 1907, p. 1. Smith, J. W.; "Effect of Weather on Yield of Corn," Monthly Weather Review, vol. 42, p. 72; and "Effect of Weather on Yield of Potatoes," ibid., vol. 43, p. 232. Brown, Wm.: "Essentials of Mental Measurement," 1911.

[2] For detailed account of the methods by which these data were obtained and the costs computed, see Report 111, Office of the Secretary, 1916.

70070°—17

TABLE I.—*Data on cost of producing baby beef.*

Farm No.	Profit per head.[1]	Weight per head.	Value per hundredweight.	Total value of feed per head.	Cost per head at weaning time.	Date of sale (months).[2]
		Pounds.				
1	+$12.07	785	$8.35	$31.49	$23.12	8.7
2	− 22.98	750	7.75	29.62	46.33	6.3
3	+ 2.79	690	7.20	20.90	30.94	2.6
4	+ 6.07	820	8.00	30.00	30.02	6.4
5	− 14.05	852	7.52	31.88	44.15	12.7
6	− 9.93	1,000	9.75	37.47	64.34	12.3
7	+ 13.68	825	8.50	32.68	25.76	8.8
8	+ 15.15	825	8.50	23.04	32.64	8.8
9	+ 27.42	800	9.50	31.06	18.01	8.5
10	− 8.92	875	9.75	59.10	33.92	8.8
11	− 19.09	922	10.14	70.52	40.20	9.9
12	+ 18.75	810	9.30	30.25	28.08	9.0
13	− 7.07	1,080	9.75	71.01	38.62	8.9
14	+ 13.53	1,048	10.11	47.43	39.83	10.0
15	+ 38.15	1,012	10.35	49.47	20.47	10.9
16	+ 9.83	1,000	9.75	40.56	42.89	10.0
17	+ 19.05	807	9.70	33.58	28.43	4.3
18	− 5.73	915	9.10	38.41	51.36	8.0
19	− 2.39	910	9.15	45.48	42.17	5.0
20	+ 10.93	890	9.70	48.95	25.86	8.2
21	+ 9.67	876	9.40	23.91	43.98	8.7
22	− 3.65	988	8.75	43.95	38.52	12.2
23	+ 49.37	1,050	9.75	27.08	27.74	10.3
24	− 7.28	798	8.25	42.84	30.09	8.0
25	− 43.00	675	8.00	44.71	50.80	6.6
26	+ 5.65	689	7.75	23.39	24.61	5.5
27	+ 0.59	860	10.00	49.74	35.85	7.5
28	− 7.71	746	8.00	26.53	40.30	6.7
29	− 2.78	850	8.90	33.35	43.46	6.5
30	− 1.36	890	7.30	20.33	49.80	3.5
31	+ 6.76	859	8.15	27.73	30.89	4.0
32	+ 7.91	765	8.10	21.02	30.83	6.0
33	+ 18.07	744	9.48	34.95	19.51	7.2
34	+ 1.33	700	9.00	31.61	28.25	5.0
35	− 32.63	740	9.25	45.72	49.76	5.9
36	+ 12.97	800	9.00	31.00	26.57	6.5
37	+ 11.15	800	7.30	17.96	27.11	2.7
38	− 43.71	740	8.50	20.12	85.66	5.0
39	+ 18.15	700	7.75	9.38	24.78	3.8
40	− 9.86	785	8.25	33.70	38.73	8.8
41	+ 2.18	656	8.14	21.90	30.09	4.9
42	+ 23.99	925	8.60	26.75	28.55	5.9
43	− 22.97	766	8.40	54.46	35.09	6.9
44	− 12.73	750	8.50	18.02	54.33	7.0
45	+ 11.80	805	8.00	20.76	34.62	5.0
46	− 22.90	924	9.85	57.59	56.63	8.0
47	+ 0.27	800	9.25	48.53	28.97	7.6
48	+ 5.37	800	8.25	29.79	29.03	5.7
49	+ 5.33	862	10.20	49.77	30.90	10.7
50	+ 2.82	800	9.00	24.81	42.86	7.0
51	+ 16.68	840	10.00	45.54	19.54	8.0
52	− 7.07	840	9.25	41.09	37.53	7.7
53	− 3.09	650	7.50	12.28	41.66	3.0
54	− 24.04	775	8.40	39.86	47.90	6.0
55	− 10.45	741	8.50	29.04	45.68	6.0
56	+ 2.83	768	7.20	24.53	29.62	6.0
57	− 0.09	1,060	8.30	47.10	45.41	3.0
58	+ 3.88	900	8.95	37.02	42.27	6.0
59	− 6.42	793	9.60	56.46	27.69	8.5
60	− 8.37	855	8.55	42.40	40.51	5.5
61	− 27.61	850	8.05	43.83	51.29	6.0
62	+ 1.64	915	8.55	34.41	39.39	6.2
63	− 0.18	811	8.60	37.13	32.90	8.9
64	− 1.00	775	8.20	18.30	43.72	5.0
65	+ 8.00	742	8.25	17.39	34.85	6.9
66	+ 5.55	827	8.50	22.30	42.05	7.3
67	+ 21.73	950	9.50	33.64	32.09	12.0
Average..	+ 0.78	834	8.76	35.02	37.01	7.2

[1] A plus sign before the quantity in this column indicates a gain; a minus sign, a loss.

[2] In order to facilitate computation, the dates have been expressed in months and decimals of a month after Jan. 1; i. e., 8.7 indicates Aug. 20, 21, or 22; 6.3 indicates June 8, 9, or 10, etc.

If, in two series of associated variables, as, say, the profit per head and the weight per head in the data under consideration, there is a tendency for a high value of the first to be associated with a high value of the second, the variables are said to be correlated, and the correlation is positive; while if a high value of the first is associated with a low value of the second, and vice versa, the correlation is said to be negative, and the best measure yet devised of the amount of the correlation is the so-called coefficient of correlation. In Table II is shown the calculation of the coefficient of correlation between profit and weight per head.

The method is as follows:

1. Find the average value for each of the variables. Here the average profit per head is $0.78, and the average weight 834 pounds.

2. Calculate the departure of the individual values from the average. In the case of record No. 1, the departure of the profit from the average is +$11.29, and of the weight, —49 pounds.

3. Find the square root of the average of the squares of these departures. This is the so-called "standard deviation," and is a measure of dispersion or the amount of variability of each variable.

4. Find the algebraic sum of the products of each pair of individual departures, i. e., for each record, multiply the departure of the profit from the average by the departure of the weight from the average, and prefix the proper sign; then find the difference between the sum of all the plus products and the sum of all the minus products.

5. Divide this result by the number of records and the standard deviation of each of the variables in turn, prefix the proper sign, and the figure obtained is the coefficient of correlation between the two factors under consideration.

If there are approximately the same number of positive and negative products and they are of the same size, it will be evident that there is no correlation, and this will be shown by the fact that the coefficient of correlation will be zero, or nearly so. If high values of the first variable are associated with high values of the second, and low values of the first with low values of the second, most of the products will be plus, and the greater their sum the closer will be the correlation and the larger will be the coefficient obtained. If a value of one variable below the average is generally associated with a value of the other above the average, the correlation will evidently be negative, and this will be shown by the fact that the sum of the products will be negative, the degree of the correlation and the size of the coefficient depending upon the size of this sum.

Expressed algebraically, the coefficient of correlation,

$$r = \frac{\Sigma xy}{n\sigma_x\sigma_y}; \qquad \text{(I)}$$

where Σxy is the sum of the products above mentioned, n is the number of pairs of variables (the same as the number of records); σ_x

TABLE II.—*Calculation of coefficient of correlation between profit and weight per head.*

Farm No.	Profit per head.[1]	x.	x^2.	Weight per head.	y.	y^2.	xy.
				Pounds.			
1	+$12.07	+11.29	+127.69	785	− 49	+2,401	− 553.7
2	− 22.98	−23.76	566.44	750	− 84	7,056	+ 1,999.2
3	+ 2.79	+ 2.01	4.00	690	−144	20,736	− 288.0
4	+ 6.07	+ 5.29	28.09	820	− 14	196	− 74.2
5	− 14.05	−14.83	219.04	852	+ 18	324	− 266.4
6	− 9.93	−10.71	114.49	1,000	+166	27,556	− 1,776.2
7	+ 13.68	+12.90	166.41	825	− 9	81	− 116.1
8	+ 15.15	+14.37	207.36	825	− 9	81	− 129.6
9	+ 27.42	+26.64	707.56	800	− 34	1,156	− 904.4
10	− 8.92	− 9.70	94.09	875	+ 41	1,681	− 397.7
11	− 19.09	−19.87	396.01	922	+ 88	7,724	− 1,751.2
12	+ 18.75	+17.97	324.00	810	− 24	576	− 432.0
13	− 7.07	− 7.85	62.41	1,080	+246	60,516	− 1,943.4
14	+ 13.53	+12.75	163.84	1,048	+214	45,796	+ 2,739.2
15	+ 38.15	+37.37	1,398.76	1,012	+178	31,684	+ 6,657.2
16	+ 9.83	+ 9.05	81.00	1,000	+166	27,556	+ 1,494.0
17	+ 19.05	+18.27	334.89	807	− 27	729	− 494.1
18	− 5.73	− 6.51	42.25	915	+ 81	6,561	− 526.5
19	− 2.39	− 3.17	10.24	910	+ 76	5,776	− 243.2
20	+ 10.93	+10.15	104.04	890	+ 56	3,136	+ 571.2
21	+ 9.67	+ 8.89	79.21	876	+ 42	1,764	+ 373.8
22	− 3.65	− 4.43	19.36	988	+154	23,716	− 677.6
23	+ 49.37	+48.59	2,361.96	1,050	+216	46,656	+10,497.6
24	− 7.28	− 8.06	65.61	798	− 36	1,296	+ 291.6
25	− 43.00	−43.78	1,918.44	675	−159	25,281	+ 6,964.2
26	+ 5.65	+ 4.87	24.01	689	−145	21,025	− 710.5
27	+ 0.59	− 0.19	.04	860	+ 26	676	− 5.2
28	− 7.71	− 8.49	72.25	746	− 88	7,744	+ 748.0
29	− 2.78	− 3.56	12.96	850	+ 16	256	− 57.6
30	− 1.36	− 2.14	4.41	890	+ 56	3,136	− 117.6
31	+ 6.76	+ 5.98	36.00	859	+ 25	625	+ 150.0
32	+ 7.91	+ 7.13	50.41	765	− 69	4,761	− 489.9
33	+ 18.07	+17.29	299.29	744	− 90	8,100	− 1,557.0
34	+ 1.33	+ 0.55	.36	700	−134	17,956	− 80.4
35	− 32.63	−33.41	1,115.56	740	− 94	8,836	+ 3,139.6
36	+ 12.97	+12.19	148.84	800	− 34	1,156	− 414.8
37	+ 11.15	+10.37	108.16	800	− 34	1,156	− 353.6
38	− 43.71	−44.49	1,980.25	740	− 94	8,836	+ 4,183.0
39	+ 18.15	+17.37	302.76	700	−134	17,956	− 2,331.6
40	− 9.86	−10.64	112.36	785	− 49	2,401	+ 519.4
41	+ 2.18	+ 1.40	1.96	656	−178	31,684	− 249.2
42	+ 23.99	+23.21	556.96	925	+ 91	8,281	+ 2,111.2
43	− 22.97	−23.75	566.44	766	− 68	4,624	+ 1,618.4
44	− 12.73	−13.63	184.96	750	− 84	7,056	+ 1,142.4
45	+ 11.80	+11.02	121.00	805	− 29	841	− 319.0
46	− 22.90	−23.68	561.69	924	+ 90	8,100	− 2,133.0
47	+ 0.27	− 0.51	.25	800	− 34	1,156	+ 17.0
48	+ 5.37	+ 4.59	21.16	800	− 34	1,156	− 156.4
49	+ 5.33	+ 4.55	21.16	862	+ 28	784	+ 128.8
50	+ 2.82	+ 2.04	4.00	800	− 34	1,156	− 68.0
51	+ 16.68	+15.90	252.81	840	+ 6	36	+ 95.4
52	− 7.07	− 7.85	60.84	840	+ 6	36	− 46.8
53	− 3.09	− 3.87	15.21	650	−184	33,856	+ 717.6
54	− 24.04	−24.82	615.04	775	− 59	3,481	+ 1,463.2
55	− 10.45	−11.23	125.44	741	− 93	8,649	+ 1,041.6
56	+ 2.83	+ 2.05	4.00	768	− 66	4,356	− 132.0
57	− 0.09	− 0.87	.81	1,060	+226	51,076	− 203.4
58	+ 3.88	+ 3.10	9.61	900	+ 66	4,356	+ 204.6
59	− 6.42	− 7.20	51.84	793	− 41	1,681	+ 295.2
60	− 8.37	− 9.15	84.64	855	+ 21	441	− 193.2
61	− 27.61	−28.36	806.56	850	+ 16	256	− 454.4
62	+ 1.64	+ 0.86	.81	915	+ 81	6,561	+ 72.9
63	− 0.18	− 0.96	1.00	811	− 23	529	+ 23.0
64	− 1.00	− 1.78	3.24	775	− 59	3,481	+ 106.2
65	+ 8.00	+ 7.22	51.84	742	− 92	8,464	− 662.4
66	+ 5.55	+ 4.77	23.04	827	− 7	49	− 33.6
67	+ 21.73	+20.95	441.00	950	+116	13,456	+ 2,436.0
	Average, +0.78		18,452.16 σ_x=$16.60	Average, 834		660.257 σ_y=99 lbs.	+30,457.6

$$r=\frac{\Sigma xy}{n\sigma_x\sigma_y}=\frac{+30457.6}{(67)(16.60)(99)}=+0.277$$

$$E_r=\pm.6745\frac{1-r^2}{\sqrt{n}}=\pm.076$$

[1] A plus sign before the quantity in this column indicates a profit, a minus sign a loss. The quantities in the column headed x are given to two places of decimals, but it was found that the use of one decimal place would give the quantities in the x^2 and xy columns with sufficient accuracy, and the computations were made accordingly. Thus, for farm No. 1, $(11.3)^2=127.69$, and $(+11.3)\ )-49)=-553.7$.

is the standard deviation of the first variable; and σ_y the standard deviation of the second. The value of "r" will always be between +1 and −1, +1 indicating perfect positive correlation, and −1 perfect negative correlation; and to be significant, the value should be appreciably greater than its probable error,

$$E_r = \frac{\pm .6745(1-r^2)}{\sqrt{n}}. \qquad \text{(II)}$$

In the example, $r=+.277$, and its probable error is $\pm.076$, so there was a tendency for the heavier calves to return a greater profit, but the correlation is by no means perfect.

PARTIAL CORRELATION.

A study in which many factors are concerned is not complete until it is determined whether or not an apparent correlation, measured in the manner explained above, is due to the fact that each of the two variables (or factors) under consideration is correlated with another or even several other variables. For instance, in the data under consideration there is apparently a high correlation between the weight of the calves and the value per hundredweight received for them ($r=+.56$), and the question now arises if heavier calves really do demand a higher price on the market. This correlation might be due entirely or in part to the fact that the heavier calves in the records obtained were sold at a later date, and that the price of cattle in general was higher later in the season; that is, the correlation exhibited here might be due to the fact that both weight and price are correlated with date of sale.

In a problem of this type, where it is necessary to consider simultaneously the relation between three variables and to determine the correlation between any two, a coefficient of net or partial correlation[1] can be determined by the formula—

$$r_{ab.c} = \frac{r_{ab} - r_{ac}r_{bc}}{\sqrt{(1-r^2_{ac})(1-r^2_{bc})}} \qquad \text{(III)}$$

Calling the three variables a, b, and c, the terms of the formula are: $r_{ab.c}$ is the coefficient of net correlation between a and b, when the effect of c is considered; r_{ab} is the ordinary coefficient of gross correlation between a and b and is obtained as explained above; r_{ac} and r_{bc} are the coefficients of gross correlation between a and c and b and c, respectively. Continuing with the example above, let us endeavor to determine the degree of correlation between weight and value per hundredweight, after taking into account any effect that date of sale might have had. In other words, we seek an answer to the question:

[1] Yule, G. U.: "Introduction to the Theory of Statistics," p. 229 et seq.

What would have been the coefficient of correlation between weight and price if all the calves had been sold on the same date? Calling the weight w, the value per hundredweight v, and the date of sale d, the gross correlation coefficients are:[1] $r_{wv}=+.56$; $r_{vd}=+.61$; $r_{wd}=+.60$. Applying the formula (III), we have:

$$r_{wv.d}=\frac{+.56-(+.61)(+.60)}{\sqrt{(1-.61^2)(1-60^2)}}=+.31$$

This value, +.31, is appreciably smaller than the value, +.56, of the gross coefficient, showing that the apparent correlation between weight and price is partly, but not entirely, due to their mutual correlation with the date of sale.

This theory can be applied to the case of several variables by a simple extension of the formula.[2]

In the general case for six variables, the total number considered in this paper—

$$r_{ab.cdef}=\frac{r_{ab.cde}-r_{af.cde}\cdot r_{bf.cde}}{\sqrt{(1-r^2_{af.cde})\,(1-r^2_{bf.cde})}}. \qquad \text{(IV)}$$

$r_{ab\cdot cdef}$ is the net coefficient of correlation between a and b, when the four factors, c, d, e, and f, are taken into account: $r_{ab\cdot cde}$, $r_{af\cdot cde}$, and $r_{bf\cdot cde}$ are the coefficients of correlation between the two variables before the period in each case when c, d, and e are taken into account.

COMPUTATION OF THE COEFFICIENTS.

The first step in the arithmetic was the computation of the gross correlation coefficients. As stated above, the variables or factors considered were: (1) The profit or loss per head; (2) weight; (3) value per hundredweight; (4) total value of feed consumed per head; (5) cost per head at weaning time; and (6) date of sale. These six variables, if taken two at a time, can be combined in 15 different ways. The first calculation was to find the coefficients of correlation between these 15 different pairs. In Table III these are the first values given. The effect of every other factor on these gross coefficients was then eliminated by successive applications of formulæ III and IV. As an example, take profit and weight, the first pair of variables correlated. The gross coefficient was first corrected for the effect of value per hundredweight, value of feed consumed, initial cost, and date of sale, in turn. Then the effect of these four factors was considered, taking them two at a time. That is, the correlation was determined when both the value per hundredweight and the cost of feed were taken into consideration at the same time. When the effect of all these factors, taking them

[1] See Table III: Correlation coefficients.

[2] Yule, G. U.: "Introduction to the Theory of Statistics," p. 229 et seq.

TABLE III.—*Correlation coefficients.*

Profit (p) and Weight (w).		Profit (p) and Value per hundredweight (v).		Profit (p) and Value of feed (f).		Profit (p) and Cost at weaning time (c).		Profit (p) and Date of Sale (d).	
r_{pw}	+0.28	r_{pv}	+0.23	r_{pf}	−0.27	r_{pc}	−0.73	r_{pd}	+0.14
$r_{pw\cdot v}$	+ .18	$r_{pv\cdot w}$	+ .10	$r_{pf\cdot w}$	− .50	$r_{pc\cdot w}$	− .78	$r_{pd\cdot w}$	− .03
$r_{pw\cdot f}$	+ .50	$r_{pv\cdot f}$	+ .56	$r_{pf\cdot v}$	− .58	$r_{pc\cdot v}$	− .73	$r_{pd\cdot v}$	.00
$r_{pw\cdot c}$	+ .48	$r_{pv\cdot c}$	+ .25	$r_{pf\cdot c}$	− .38	$r_{pc\cdot f}$	− .75	$r_{pd\cdot f}$	+ .29
$r_{pw\cdot d}$	+ .24	$r_{pv\cdot d}$	+ .19	$r_{pf\cdot d}$	− .37	$r_{pc\cdot d}$	− .73	$r_{pd\cdot c}$	+ .16
$r_{pw\cdot vf}$	+ .39	$r_{pv\cdot wf}$	+ .48	$r_{pf\cdot wv}$	− .64	$r_{pc\cdot wv}$	− .78	$r_{pd\cdot wv}$	− .08
$r_{pw\cdot vc}$	+ .43	$r_{pv\cdot wc}$	− .04	$r_{pf\cdot wc}$	− .83	$r_{pc\cdot wf}$	− .92	$r_{pd\cdot wf}$	+ .06
$r_{pw\cdot vd}$	+ .20	$r_{pv\cdot wd}$	+ .12	$r_{pf\cdot wd}$	− .50	$r_{pc\cdot wd}$	− .79	$r_{pd\cdot wc}$	− .18
$r_{pw\cdot fc}$	+ .85	$r_{pv\cdot fc}$	+ .71	$r_{pf\cdot vc}$	− .74	$r_{pc\cdot vf}$	− .83	$r_{pd\cdot vf}$	+ .02
$r_{pw\cdot fd}$	+ .43	$r_{pv\cdot fd}$	+ .50	$r_{pf\cdot vd}$	− .58	$r_{pc\cdot vd}$	− .73	$r_{pd\cdot vc}$	+ .02
$r_{pw\cdot cd}$	+ .49	$r_{pv\cdot cd}$	+ .18	$r_{pf\cdot cd}$	− .50	$r_{pc\cdot fd}$	− .77	$r_{pd\cdot fc}$	+ .38
$r_{pw\cdot vfc}$	+ .91	$r_{pv\cdot wfc}$	+ .83	$r_{pf\cdot wvc}$	− .95	$r_{pc\cdot wvf}$	− .97	$r_{pd\cdot wvf}$	− .15
$r_{pw\cdot vfd}$	+ .42	$r_{pv\cdot wfd}$	+ .49	$r_{pf\cdot wvd}$	− .65	$r_{pc\cdot wvd}$	− .79	$r_{pd\cdot wvc}$	− .19
$r_{pw\cdot vcd}$	+ .46	$r_{pv\cdot wcd}$	+ .04	$r_{pf\cdot wcd}$	− .83	$r_{pc\cdot wfd}$	− .92	$r_{pd\cdot wfc}$	− .10
$r_{pw\cdot fcd}$	+ .83	$r_{pv\cdot fcd}$	+ .65	$r_{pf\cdot vcd}$	− .74	$r_{pc\cdot vfd}$	− .83	$r_{pd\cdot vfc}$	+ .06
$r_{pw\cdot vfcd}$	+ .97	$r_{pv\cdot wfcd}$	+ .94	$r_{pf\cdot wvcd}$	− .98	$r_{pc\cdot wvfd}$	− .98	$r_{pd\cdot wvfc}$	+ .77

Weight (w) and Value per hundredweight (v).		Weight (w) and Value of feed (f).		Weight (w) and Cost at weaning time (c).		Weight (w) and Date of sale (d).		Value per hundredweight (v) and Value of feed (f).	
r_{wv}	+0.56	r_{wf}	+0.51	r_{wc}	+0.07	r_{wd}	+0.60	r_{vf}	+0.65
$r_{wv\cdot p}$	+ .53	$r_{wf\cdot p}$	+ .63	$r_{wc\cdot p}$	+ .42	$r_{wd\cdot p}$	+ .59	$r_{vf\cdot p}$	+ .76
$r_{wv\cdot f}$	+ .35	$r_{wf\cdot v}$	+ .23	$r_{wc\cdot v}$	+ .15	$r_{wd\cdot v}$	+ .39	$r_{vf\cdot w}$	+ .51
$r_{wv\cdot c}$	+ .57	$r_{wf\cdot c}$	+ .51	$r_{wc\cdot f}$	+ .08	$r_{wd\cdot f}$	+ .49	$r_{vf\cdot c}$	+ .66
$r_{wv\cdot d}$	+ .31	$r_{wf\cdot d}$	+ .36	$r_{wc\cdot d}$	+ .12	$r_{wd\cdot c}$	+ .60	$r_{vf\cdot d}$	+ .55
$r_{wv\cdot pf}$	+ .10	$r_{wf\cdot pv}$	+ .42	$r_{wc\cdot pv}$	+ .42	$r_{wd\cdot pv}$	+ .40	$r_{vf\cdot pw}$	+ .65
$r_{wv\cdot pc}$	+ .53	$r_{wf\cdot pc}$	+ .86	$r_{wc\cdot pf}$	+ .81	$r_{wd\cdot pf}$	+ .42	$r_{vf\cdot pc}$	+ .84
$r_{wv\cdot pd}$	+ .28	$r_{wf\cdot pd}$	+ .50	$r_{wc\cdot pd}$	+ .45	$r_{wd\cdot pc}$	+ .61	$r_{vf\cdot pd}$	+ .68
$r_{wv\cdot fc}$	+ .36	$r_{wf\cdot vc}$	+ .22	$r_{wc\cdot vf}$	+ .13	$r_{wd\cdot vf}$	+ .39	$r_{vf\cdot wc}$	+ .52
$r_{wv\cdot fd}$	+ .14	$r_{wf\cdot vd}$	+ .24	$r_{wc\cdot vd}$	+ .16	$r_{wd\cdot vc}$	+ .39	$r_{vf\cdot wd}$	+ .50
$r_{wv\cdot cd}$	+ .32	$r_{wf\cdot cd}$	+ .36	$r_{wc\cdot fd}$	+ .12	$r_{wd\cdot fc}$	+ .50	$r_{vf\cdot cd}$	+ .56
$r_{wv\cdot pfc}$	− .67	$r_{wf\cdot pvc}$	+ .88	$r_{wc\cdot pvf}$	+ .89	$r_{wd\cdot pvf}$	+ .42	$r_{vf\cdot pwc}$	+ .88
$r_{wv\cdot pfd}$	− .09	$r_{wf\cdot pvd}$	+ .44	$r_{wc\cdot pvd}$	+ .45	$r_{wd\cdot pvc}$	+ .43	$r_{vf\cdot pwd}$	+ .65
$r_{wv\cdot pcd}$	+ .27	$r_{wf\cdot pcd}$	+ .80	$r_{wc\cdot pfd}$	+ .79	$r_{wd\cdot pfc}$	+ .36	$r_{vf\cdot pcd}$	+ .77
$r_{wv\cdot fcd}$	+ .16	$r_{wf\cdot vcd}$	+ .22	$r_{wc\cdot vfd}$	+ .14	$r_{wd\cdot vfc}$	+ .40	$r_{vf\cdot wcd}$	+ .50
$r_{wv\cdot pfcd}$	− .90	$r_{wf\cdot pvcd}$	+ .97	$r_{wc\cdot pvfd}$	+ .96	$r_{wd\cdot pvfc}$	+ .81	$r_{vf\cdot pwcd}$	+ .97

Value per hundredweight (v) and Cost at weaning time (c).		Value per hundredweight (v) and Date of sale (d).		Value of feed (f) and Cost at weaning time (c).		Value of feed (f) and Date of sale (d).		Cost at weaning time (c) and Date of sale (d).	
r_{vc}	−0.09	r_{vd}	+0.61	r_{fc}	+0.01	r_{fd}	+0 42	r_{cd}	−0.04
$r_{vc\cdot p}$	+ .12	$r_{vd\cdot p}$	+ .60	$r_{fc\cdot p}$	− .28	$r_{fd\cdot p}$	+ .48	$r_{cd\cdot p}$	+ .09
$r_{vc\cdot w}$	− .16	$r_{vd\cdot w}$	+ .41	$r_{fc\cdot w}$	− .03	$r_{fd\cdot w}$	+ .16	$r_{cd\cdot w}$	− .11
$r_{vc\cdot f}$	− .13	$r_{vd\cdot f}$	+ .49	$r_{fc\cdot v}$	+ .10	$r_{fd\cdot v}$	+ .03	$r_{cd\cdot v}$	+ .02
$r_{vc\cdot d}$	− .08	$r_{vd\cdot c}$	+ .61	$r_{fc\cdot d}$	+ .04	$r_{fd\cdot c}$	+ .42	$r_{cd\cdot f}$	− .05
$r_{vc\cdot pw}$	− .14	$r_{vd\cdot pw}$	+ .42	$r_{fc\cdot pw}$	− .79	$r_{fd\cdot pw}$	+ .17	$r_{cd\cdot pw}$	− .21
$r_{vc\cdot pf}$	+ .54	$r_{vd\cdot pf}$	+ .41	$r_{fc\cdot pv}$	− .58	$r_{fd\cdot pv}$	+ .04	$r_{cd\cdot pv}$	+ .03
$r_{vc\cdot pd}$	+ .08	$r_{vd\cdot pc}$	+ .59	$r_{fc\cdot pd}$	− .37	$r_{fd\cdot pc}$	+ .53	$r_{cd\cdot pf}$	+ .27
$r_{vc\cdot wf}$	− .17	$r_{vd\cdot wf}$	+ .39	$r_{fc\cdot wv}$	+ .07	$r_{fd\cdot wv}$	− .07	$r_{cd\cdot wv}$	− .05
$r_{vc\cdot wd}$	− .13	$r_{vd\cdot wc}$	+ .40	$r_{fc\cdot wd}$	− .01	$r_{fd\cdot wc}$	+ .16	$r_{cd\cdot wf}$	− .10
$r_{vc\cdot fd}$	− .12	$r_{vd\cdot fc}$	+ .48	$r_{fc\cdot vd}$	+ .10	$r_{fd\cdot vc}$	+ .03	$r_{cd\cdot vf}$	+ .01
$r_{vc\cdot pwf}$	+ .77	$r_{vd\cdot pwf}$	+ .41	$r_{fc\cdot pwv}$	− .92	$r_{fd\cdot pwv}$	− .15	$r_{cd\cdot pwv}$	− .18
$r_{vc\cdot pwd}$	− .05	$r_{vd\cdot pwc}$	+ .40	$r_{fc\cdot pwd}$	− .77	$r_{fd\cdot pwc}$	+ .01	$r_{cd\cdot pwf}$	− .13
$r_{vc\cdot pfd}$	+ .49	$r_{vd\cdot pfc}$	+ .33	$r_{fc\cdot pvd}$	− .58	$r_{fd\cdot pvc}$	+ .07	$r_{cd\cdot pvf}$	+ .06
$r_{vc\cdot wfd}$	− .14	$r_{vd\cdot wfc}$	+ .38	$r_{fc\cdot wvd}$	+ .06	$r_{fd\cdot wvc}$	− .06	$r_{cd\cdot wvf}$	− .04
$r_{vc\cdot pwfd}$	+ .90	$r_{vd\cdot pwfc}$	+ .81	$r_{fc\cdot pwvd}$	− .96	$r_{fd\cdot pwvc}$	− .82	$r_{cd\cdot pwvf}$	− .75

two at a time, had been considered, they were taken three at a time, and finally all four were taken into account simultaneously. In all, 260 of the coefficients were computed. Some of them, however, are of little interest, and if this were simply a study of the data, and not also an exposition of the method used, the computation of some of them might have been omitted.

In order to avoid needless repetition in the tables, the different factors are designated by letters as follows: Profit=p, weight=w, value per hundredweight=v, value of feed per head=f, cost at weaning time=c, date of sale=d, and the notation for the different coefficients is the same as that used in the explanation of the theory, viz: $r_{ab \cdot cd, \text{ etc.}}$, is the coefficient of correlation between a and b when c, d, etc., are taken into account.

INTERPRETATION OF THE COEFFICIENTS.

There are four factors, namely, initial cost, value of feed consumed, weight, and selling price, which determine almost entirely the profit or loss to the farmer in finishing cattle for market. In fattening baby beef animals, the weight of the calves and the value of feed consumed both depend, to a large extent, on their age when sold and the length of time they were on feed. Also the price per pound received for them is rather intimately connected with the date on which they were sold, prices having had a tendency to rise as the season advanced. The calves for which data were gathered were all born in the spring, went on feed in the fall or early winter, and were sold some time during the following year. Consequently, any one of the three factors, age, length of feeding period, and date of sale, is a very good measure of the other two, and on account of this only the date of sale has been considered.

If the price per pound, value of feed, initial cost, and date of sale were constant, and if nothing else affected the profit, it would vary directly with the weight in every case and, according to the theory, the coefficient of correlation between the two should be +1. The coefficient, $r_{pw \cdot vfcd}$, obtained here is +.97. Similarly, if all things were constant except value per hundredweight and profit, there would be perfect positive correlation between them. The net coefficient, $r_{pv \cdot wfcd}$, given in Table III is +.94. If all things were constant except the value of feed consumed we should expect a high negative correlation between it and profit, i. e., the calf that received the least feed would return the greatest profit. The net coefficient, $r_{pf \cdot wvcd}$, is —.98. Similarly, other things being equal, perfect negative correlation should exist between initial cost and profit. The net coefficient in this case, $r_{pc \cdot wvfd}$, is also —.98. An examination of the remainder of these net coefficients, which are the last ones given in the table,

will show that, with the exception of the five between date of sale and the other variables, they are all numerically equal to or above .90. It has been shown that part, but not all, of the correlation between weight and price was due to the date of sale, and since date of sale is only an approximate measure of age and length of feeding period, it would not be reasonable to expect the net correlation between it and the other variables to be perfect. The fact that all the net coefficients except these five are so nearly +1 or −1, when there was every reason to expect perfect correlation, is striking proof of the reliability of this method of analysis as well as of the accuracy of data such as those under consideration, and is at the same time a very good check on the computations.

In the interpretation of the coefficients care must be taken to distinguish between subjective and relative factors, i. e., between cause and effect. Most interest is naturally attached to determining to just what extent each of the factors under consideration is responsible for the farmer's loss or gain in his baby-beef enterprise, and here there can be no confusion of cause and effect, for all the other factors are necessarily causative. Throughout the remainder of the investigation the amount of profit or loss is an effect and not a cause, and consequently too much weight should not be given to a coefficient in which the effect of profit has been taken into account.

THE APPARENT CORRELATIONS.

In taking up the discussion of the coefficients themselves, the apparent correlations between profit and the other five factors are first considered:

Coefficients of correlation.

Profit and Weight.	Profit and Value per hundredweight.	Profit and Value of feed.	Profit and Cost at weaning time.	Profit and Date of sale.
+.28	+.23	−.27	−.73	+.14

These five coefficients should show the average effect of each of the five factors on the profit. The coefficient for profit and date of sale (+.14) shows that the profit on the calves sold early in the season was practically as great as on those sold later. The first three are all of nearly the same size, but are too small to indicate more than slight relationship. In regard to them we may say, therefore, that in the data under consideration: (1) There was a tendency for the heavier calves to return a greater profit; (2) there is some correlation between price per pound and profit; (3) generally speaking, the farmer whose calves consumed feed worth more than the average made a profit somewhat less than the average.

A very high degree of correlation between profit and cost at weaning time is shown by the coefficient —.73, and, as would be expected, it is negative. The size of this coefficient as compared to the others indicates that the cost of producing the calves and carrying them until weaning time is by far the most important factor in determining the profit derived by any particular farmer from the production of baby beef. In all of the records considered the calves were with the cows until they went on feed, and there was no expense directly chargeable to them. Bearing this in mind, the further statement is justified that the cost of maintaining the breeding herd and the size of the calf crop have considerably more to do with the profitableness of the enterprise than the actual preparation of the calves for market.

Coefficients of correlation between weight and factors other than profit.

Weight and Value per hundredweight.	Weight and Value of feed.	Weight and Cost at weaning time.	Weight and Date of sale.
+.56	+.51	+.07	+.60

The coefficient +.07, for weight and cost at weaning time, is the most striking one given here. Its very small size shows that there is no connection between the cost of the calves up to the time they went on feed and the weights at which they were sold. The cost of a calf at weaning time is determined very largely by the manner in which the breeding herd is handled, and consequently this coefficient shows further that on the farms studied the calves from the herds which were maintained at a low cost per head weighed just as much when sold as did those from herds having a high maintenance cost. The coefficients for weight and value of feed and weight and date of sale are what should normally be expected. The calves that received more feed than the average weighed more than the average, and the ones that were sold in the latter part of the season also weighed more than the average. The high correlation, exhibited by the coefficient +.56, between weight and price per pound is a surprising one, but it will be shown later that it is almost entirely due to the mutual correlation of these two factors with some of the others.

The gross coefficient for value per hundredweight and value of feed,[1] +.65, shows another apparently high correlation which may or may not disappear when some of the other factors are taken into account. There is no correlation between value per hundredweight and cost at weaning time. The correlation between value per pound and date of sale is shown by the coefficient +.61, which confirms

[1] For this and all coefficients mentioned hereafter, see Table III.

the statement already made that the price was generally higher later in the season. The remaining gross coefficients are +.01 for total value of feed consumed per head and cost at weaning time, +.42 for value of feed consumed and date of sale, and —.04 for cost at weaning time and date of sale. The coefficients +.01 and —.04 show that cost at weaning time is uncorrelated with either value of feed consumed or date of sale. With regard to the correlation between value of feed consumed per head and date of sale, we may say that the value of feed consumed is probably very nearly proportional to the length of the feeding period, and if the actual length of time on feed had been used here instead of its approximate measure, the date of sale, the correlation would probably have been higher.

EFFECT OF THE OTHER FACTORS ON THE APPARENT CORRELATIONS.

The small degree of correlation present between profit and weight is mostly due to differences in price, the coefficient being reduced from +.28 to +.18, when the value per hundredweight is taken into account; that is to say, the tendency of the heavier calves to be the more profitable is mostly due to the fact that they sold for a better price per pound than that commanded by the smaller calves.

The coefficient $r_{pw.f}$ is +.50, which is considerably higher than the gross coefficient, showing that if the value of feed had been constant while other things remained unchanged, the correlation between profit and weight would have been greater.

The correct explanation of the size of the coefficient $r_{pw.c}$, which is +.48, is not so apparent. It indicates, however, that if the influence of the cost at weaning time, the factor most closely related to profit, were eliminated, the correlation between profit and weight would be greater.

When the date of sale is taken into account, the correlation between profit and weight becomes somewhat less than the gross correlation, but the difference is not enough to be significant.

The coefficients obtained for the correlation between weight and profit, when the effect of the other factors, two at a time, is considered, are generally higher than when they are considered one at a time. This means that if the influence of two of the factors contributing to the profit or loss is eliminated, its correlation with any of the remaining factors is higher than if the influence of but one had been eliminated.

It is interesting to note here that the correlation between weight and profit, even when the other factors are taken into account, is almost entirely independent of the date of sale. The apparent correlation, +.28, becomes +.24 when date of sale is taken into account. When value per pound is taken into account, the coefficient is +.18; when price and date of sale are considered simultaneously,

the coefficient is +.20. Similarly, $r_{pw.f} = +.50$, and $r_{pw.fd} = +.43$; $r_{pw.c} = +.48$, and $r_{pw.cd} = +.49$; $r_{pw.vf} = +.39$, and $r_{pw.vfd} = +.42$; $r_{pw.vc} = +.43$, and $r_{pw.vcd} = +.46$; $r_{pw.fc} = +.85$, and $r_{pw.fcd} = +.83$; $r_{pw.vfc} = +.91$, and $r_{pw.vfcd} = +.97$.

The remainder of the coefficients will not be taken up in detail, for the same reasoning may be applied as has been used for those between profit and weight. The notation is consistent throughout, and the arrangement is such that any desired coefficient can be found.

There does not seem to be any relation between cost at weaning time and any of the other factors considered, except profit, and since cost at weaning time had more influence on profit than any of the others, it might be of interest to know the relationship that would have existed between profit and the other factors if the initial cost had been constant.

The coefficients are as follows:

$r_{pw.c.}$	$r_{pv.c.}$	$r_{pf.c.}$	$r_{pd.c.}$
+.48	+.25	−.38	+.16

From these coefficients, it is evident that if the initial cost of all the calves had been the same, the most important factor in determining the profit would have been the weight when marketed; the other factors in the order of their importance being the total value of feed consumed, the price per pound, and the date of sale. However, the correlation between profit and date of sale is still too small to be important.

The statement has already been made that the apparent correlation between weight and value per hundredweight ($r = +.56$) is due to the effect of other factors. A study of the coefficients obtained when these other factors are taken into consideration shows that when the influence of date of sale is eliminated, the coefficient is reduced to +.31; when the influence of the value of feed consumed is eliminated, the coefficient becomes +.35; and when the two factors are taken into account simultaneously, the coefficient is +.14. This shows that the quantity of feed consumed per head was responsible for nearly as much of this correlation as was the date of sale, and that the two together account for practically the whole of it. In other words, the value of feed consumed and the date of sale need to be considered simultaneously here, because the later the date of sale, the longer is the feeding period, and consequently the greater the quantity and value of feed consumed.

The gross correlation between date of sale and value per pound is shown by the coefficient +.61, and that between total value of feed consumed per head and value per pound, by the coefficient +.65.

These rather large coefficients become very little smaller when all the other causal factors are taken into account. Therefore, there must be some relationship existing between value per pound and date of sale, and value per pound and value of feed consumed. The reason for the correlation between value per pound and date of sale has already been given. It is probable that the reason for the high correlation between the value per pound and value of feed consumed is due to the fact that the calves which were fed the heaviest ration, regardless of the length of feeding period, were the fattest when marketed, and consequently sold at a higher price. However, the relation between the profit and value of feed consumed per head as measured by the correlation coefficient r_{pf} is —.27, and when the influence of a longer feeding period is taken into account by eliminating the effect of date of sale the correlation is still negative ($r_{pf \cdot d} = -.37$).

SUMMARY.

The results show that data such as those obtained by farm management surveys can be analyzed very thoroughly by the use of the correlation coefficients. It is generally known before the analysis is attempted which factors are causal and which resultant, and consequently there should be very little difficulty in interpreting the coefficients correctly. The coefficients of net correlation afford a very good means of determining the net effect of each of several factors bearing upon a result, or of eliminating the effect of other factors when it is desired to find the true relationship existing between any two. Although it is not possible to give a definite concrete meaning to correlation coefficients, they are very concise relative measures of the degree of relationship existing between the factors being studied. They therefore give the investigator a single index which will show what, by the ordinary tabular method, it takes a whole table to show. While properly constructed tables will show whether or not any relationship exists between two factors, it is a difficult matter to determine which of two causes, say, has the greater effect on the result, and it is impossible, without a large number of records and a great amount of sorting and tabulation, to separate all the factors being considered in a study and find the effect that each one would have had if the others had not been present, or if they had been constant throughout the investigation. If the gross coefficients of correlation between every pair of factors have been determined, it is possible to find these relationships by simply substituting in the formula for determining a net coefficient from the gross coefficients, without any further reference to the records themselves. This method should be especially use-

ful if only a limited number of records or observations are available, for it does away with the necessity of sorting into many groups, with the consequent falling off in the reliability of the averages obtained.

The analysis of the data on fattening baby beef animals showed:

(1) That for the herds considered, the cost of producing the calves and carrying them until weaning time was by far the most important factor in determining the profit;

(2) That there was no connection between the cost at weaning time and any of the other factors, for the calves which were produced cheaply were seemingly just as good feeders and brought just as good a price per pound as the more expensive ones;

(3) That the weight at which the calves were sold and the date of sale had very little effect on the profit, except for the fact that in the two years of the records the price was higher in the latter part of the summer, at the time when the heavier calves were put on the market;

(4) That the calves which consumed the heaviest ration sold at higher prices than the others, but did not return a correspondingly greater profit, as the advanced price scarcely offset the extra value of feed consumed.

OTHER PUBLICATIONS OF THE UNITED STATES DEPARTMENT OF AGRICULTURE RELATING TO THE SUBJECT OF THIS BULLETIN.

Elementary Notes on Least Squares, the Theory of Statistics and Correlation, for Meteorology and Agriculture. (Monthly Weather Review, vol. 44, 1916, p. 551.)

Effect of Weather on Yield of Potatoes. (Monthly Weather Review, vol. 43, 1915, p. 232.)

Effect of Weather on Yield of Corn. (Monthly Weather Review, vol. 42, 1914, p. 72.)

Methods and Cost of Growing Beef Cattle in the Corn Belt States. (Report No. 111, Office of the Secretary.)

ADDITIONAL COPIES

OF THIS PUBLICATION MAY BE PROCURED FROM
THE SUPERINTENDENT OF DOCUMENTS
GOVERNMENT PRINTING OFFICE
WASHINGTON, D. C.
AT

5 CENTS PER COPY